DOCUMENT RESUME

ED 315 440 TM 014 463

AUTHOR	Campo, Stephanie F.
TITLE	Canonical Correlation as the Most General Parametric Method: Implications for Educational Research.
PUB DATE	Jan 90
NOTE	28p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 25-27, 1990).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Computer Software; *Educational Research; *Heuristics; *Multivariate Analysis
IDENTIFIERS	*Parametric Analysis; Statistical Analysis System; Variance (Statistical)

ABSTRACT

All parametric methods are special cases of canonical correlation analysis and can, in fact, be performed using the canonical correlation procedures found in commonly available computer statistics packages such as the Statistical Analysis System (SAS) software. It is suggested that canonical analysis has advantages of versatility since it can accommodate various levels of scale and does not require that variance be discarded or that variable relationships be distorted. In addition, it has advantages of scientific parsimony since the analysis is also a variable reduction technique not unlike principal components analysis. Finally, canonical correlation analysis has heuristic value since it provides a framework that allows educational researchers to understand how all parametric methods are related. A small data set and the accompanying SAS program are presented, and it is established through concrete demonstration that canonical analysis does subsume other methods as special cases. Fourteen data tables and sample SAS commands are included. (Author/TJH)

canon90.wp 1/18/90

Canonical Correlation as the Most General Parametric Method: Implications for Educational Research

Stephanie F. Campo

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 27, 1990.

Abstract

It is noted in the present paper that all parametric methods are special cases of canonical correlation analysis, and can in fact be performed using the canonical correlation procedures in commonly available computer statistics packages such as SAS. It is suggested that canonical analysis has (a) advantages of versatility, since the method can accomodate various levels of scale and does not require that variance be discarded or that variable relationships be distorted, (b) advantages of scientific parsimony, since the analysis is also a variable reduction technique not unlike principal components analysis, and (c) heuristic value, since the analytic method provides a framework to allow educational researchers to understand how all parametric methods are related. A small data set and the accompanying SAS program is presented, and it is established through concrete demonstration that canonical analysis does subsume other methods as special cases.

Canonical Correlation as the Most
General Parametric Method

Educational problems are complex and necessarily involve the understanding of multiple variables and their interrelationships, so that educational problems will be realistically portrayed in statistical analyses. Multivariate statistical methods deal with the full system of interrelationships between variables and thus shed the most light upon how variables work together and influence each other.

Univariate statistical techniques are not designed to take into account the relationships existing simultaneously between multiple variables. Thus, the use of univariate techniques in educational research inquiry can lead to seriously spurious results and interpretations (Fish, 1988; Maxwell, in press).

Canonical correlation analysis is a multivariate method which facilitates understanding of the relationships between multiple variable sets, each of which consist of two or more variables (Thompson, 1984). Since canonical correlation analyses can investigate multiple independent and dependent variables, the complex relationships among variables may be discovered as they exist within the **actual complex reality** of the educational situation which we wish to understand.

Ever since Cohen's (1968) classic article appeared, more and more researchers have come to appreciate that regression as a general linear model subsumes all univariate parametric methods as

special cases. This realization led to increased use of regression methods for various research problems. Knapp (1978) extended Cohen's argument by showing mathematically that canonical correlation analysis is the most general case of the general linear model.

The view of canonical analysis as the most general linear model is one that is important to convey in instructional settings. The insight provides a framework within which students can relate the various parametric methods to each other. Furthermore, the realization suggests that canonical analysis may be useful in many research situations. Thompson (1989) presents a bibliography of canonical applications illustrating the power of the method.

The purpose of the present paper is to illustrate in a very concrete and non-mathematical manner how canonical analysis yields results identical to those achieved using other parametric methods. A data set is analyzed using canonical analysis and *t*-tests, canonical analysis and ANOVA, and so forth, and it is shown how the results are equivalent. This makes the realization that canonical analysis subsumes parametric methods hands-on and not an abstract mathematical argument. The code from a Statistical Analysis System (SAS) computer run used to generate these results is appended for persons who might wish to replicate the results reported here.

Comparison of Canonical Correlation and Principal Components

Canonical correlation analysis is a variable reduction technique not unlike the more familiar principal component method

used by many researchers in a factor analytic context (Stevens, 1986). Both methods aim to reduce the number of variables in a variable set to synthetic or latent variables that represent all or most of the variance among variables (Stevens, 1986). The principal component method accounts for the maximum amount of correlation within a single set of variables while canonical correlation analysis accounts for the maximum amount of correlation between two variable sets, i.e., predictor or independent variables and criterion or dependent variables (Chacko, 1986). Thompson (1984) presents the calculation of the so-called "quadruple product matrix" that is actually subjected to a principal components analysis in canonical correlation investigations--the details of this mathematical computation are beyond the scope of this presentation.

Canonical correlation analysis focuses on the matrix of bivariate correlation coefficients involving <u>all</u> the variables in a given study. These variables may be of any level of scale. Canonical correlation analysis deals with the shared variance existing between variables and does not (a) discard variance by restricting the level of scale of the independent variable(s) to the nominal level, nor does canonical analysis (b) distort the actual relationships among the variables which exist in reality, as do ANOVA and MANOVA with respect to independent variables (Thompson, 1984).

Thompson (1988a) states that "Even when analysis of variance methods represent good analytic choices, regression or general

linear model approaches to the methods using *a priori* contrast coding still tend to be suprrior since these approaches tend to yield greater power against Type II error and tend to be more theoretically grounded" (p. 2). Thompson (1988b) points out that the incorrect use of OVA methods in analyzing results occurs because some researchers incorrectly associate the use of OVA methods of analysis with the use of experimental designs. But it is the design and *not* the analytic method that yields power to make causal inferences. More detail on these issues is available in excellent treatments by Benton (1990) and by Keppel and Zedeck (1989).

Problems with Conducting Multiple Univariate Tests

Employing several separate univariate tests (e.g., *t*-tests, ANOVA, chi square contingency table tests) within a single study is problematic on three accounts. First, the univariate tests fail to account for the shared variance between dependent variables and multiple independent variables, which exists in nature, thereby distorting the reality to which the researcher purportedly hopes to generalize (Thompson, 1988a). Second, the experimentwise alpha level, which is the probability of one or more Type I errors in the study **as a whole**, accumulates or inflates from employing multiple hypothesis tests in a single study (Thompson, 1988a). Where *k* is the number of separate tests in the study and where each test is conducted at the same alpha level, the probability of a Type I error in one or more of a set of several hypothesis tests

can inflate to as high as:

$$1 - (1 - \text{alpha})^k.$$

Last, the researcher employing several separate univariate tests, rather than using a multivariate test, may fail to find statistically significant results which may have been found using multivariate analysis. Thompson (1986), Fish (1988), and Maxwell (in press) each provide example data analyses illustrating that multivariate tests can isolate statistically significant results for cases in which multiple univariate tests were unable to detect statistical significance.

Thompson (1988a) concludes that univariate tests are often not adequate to deal with the shared variance among independent and dependent variables, and can therefore lead to both Type I and Type II errors. Huberty and Morris (1989) concur. One important multivariate method that avoids these problems is canonical correlation analysis.

Canonical Correlation as the Most General Parametric Method

Equating canonical and other parametric results. Thompson (1988a) has demonstrated that canonical correlation analysis can be used to yield results exactly identical to those from both conventional univariate and multivariate analyses, thus showing that canonical correlation analysis, not multiple regression, is the most general case of the general linear model. A priori contrast coding of data must be performed in order to equate

canonical results with univariate and multivariate analyses involving group membership, e.g., *t*-tests and OVA analyses.

Advantages of the canonical method. The comparison of univariate and multivariate methods to the canonical results reveals at least three important advantages of the canonical analytic method. First, canonical analysis allows the most **versatility** for the analysis of independent and dependent variables, because in canonical analysis the independent and dependent variables may be of any level of scale and multiple independent and dependent variables may be analyzed.

Recall that the *t*-test, ANOVA, and multiple regression methods analyze a single, intervally scaled dependent variable. Further, the *t*-test, ANOVA, and MANOVA methods restrict the independent variable to the nominal level of scale. A single dependent variable at the nominal level of scale and multiple independent variables at any level of scale may be analyzed using discriminant analysis (Huberty & Wisenbaker, in press). Variables may be scaled at any level when estimating some variation of the bivariate correlation coefficient.

Second, canonical correlation analysis serves the cause of scientific **parsimony** because canonical correlation analysis reduces the complex relationships between the set of independent and dependent variables through the creation of a **smaller number** of uncorrelated pairs of linear combinations of variables that best express the correlations between the two sets of variables (Afifi & Clark, 1984). The linear combinations are called "canonical

variables" (Afifi & Clark, 1984) or "canonical functions" (Thompson, 1984). The Pearson correlation between corresponding pairs of canonical variate scores are called canonical correlation coefficients, or R_c's.

Stevens (1986) illustrates an example of the parsimony of canonical description when he proposes that if a researcher were interested in understanding the nature of relationships between variable sets X and Y, where set X contains 5 variables and set Y contains 10 variables, there would be 50 correlations between variables for the researcher to consider. The canonical analysis, however, will reduce the number of correlations that the researcher needs to consider while still accounting for the maximum amount of correlation between the two sets of variables. In this example, there will be five canonical correlations, because the number of R_c's equals the number of variables in the smaller variable set (Thompson, 1984). Each canonical function represents the highest intercorrelations possible between the two sets of variables, subject to the restriction that none of the functions are correlated with each other (Chacko, 1986).

Stevens (1986) describes the successive steps in the canonical procedure. Step one involves finding two linear combinations, one from the set of independent variables and the other from the set of dependent variables, which have the maximum possible Pearson correlation. The largest Pearson product-moment correlation between the scores on the two linear composites (i.e., the linear combinations from the independent and dependent variable sets)

represents the largest canonical correlation, R_c.

Step two involves finding a second pair of linear combinations which are perfectly uncorrelated with the first pair of linear combinations, and which have the maximum possible correlation given this restriction. The correlation between the scores on these two linear composites represents the second largest canonical correlation. The canonical procedure continues in this manner until all possible R_c's are computed, with each successive linear combination accounting for the maximum amount of variance that is left between the variables.

The third advantage of canonical analysis is that canonical analysis has value as a **heuristic framework in instruction**--teaching canonical concepts can provide students with important insights regarding the relatedness of all parametric methods. These benefits are revealed by a comparison of selected univariate and multivariate methods to results from canonical analyses. An important insight is that all parametric methods are special cases of canonical analysis, and can be performed with canonical analysis, and that the reverse is not true, and simpler parametric methods cannot be used to conduct canonical analyses.

It is also important for students to see that all parametric methods involve the creation of "synthetic" scores for each person and that these "synthetic" scores become the focus for analysis, just as in all parametric methods (Thompson, 1988a). For example, in multiple regression, beta weights are applied to the participants' raw scores to yield synthetic or latent scores

sometimes called "YHAT" (Thompson, 1988a). The correlation between participants' actual dependent variable scores and the synthetic dependent variable scores (YHAT) is the multiple correlation coefficient. Similar synthetic variables are created using weights (e.g., beta weights, function coefficients, factor pattern coefficients) in all parametric analyses.

Finally, it is also important for students to see that <u>all parametric methods involve effect size estimates</u>. Researchers routinely interpret r^2, R^2, and Rc^2 in correlational studies. Equivalent results (e.g., eta^2 or the correlation ratio, $omega^2$) with ANOVA and related analyses, and it is equally important that such effect sizes be consulted during interpretation (Thompson, 1988b).

<u>Presentation of Canonical Analysis to Yield Results Identical to Other Univariate and Multivariate Tests</u>

The Statistical Analysis System (SAS) program presented in Appendix A was used to analyze the data presented by McMurray (1987). Table 1 contains the data for 52 participants who rated the helpfulness of ten instructional strategies using a Likert scale with intervals ranging from 1 to 10. The total score on this instrument was selected as a dependent or criterion variable and received the variable label SURVEY. Data for a second dependent variable was created and the variable label CLASSAVG (i.e., score in the course) was given. The independent or predictor variables (and variable names) were SEX, AGE, STATUS (freshman or sophomore),

and GPA (grade point average).

INSERT TABLE 1 ABOUT HERE

Table 1 also contains the contrast coding columns for variables SEX, STATUS, and the interaction of SEX and STATUS. Variables SX1 and ST1 express the information involved in the variables SEX and STATUS respectively. Variable SX1ST1 expresses the interaction of SEX and STATUS. The contrast coding columns represent orthogonal coding, meaning that the correlation between the contrast variables is exactly zero.

Table 2 presents the Pearson product-moment correlation between the variables AGE and CLASSAVG. The correlation coefficient was computed to be -.28910 with an associated *p* calculated value of .0376. A canonical correlation analysis involving the same two variables (see Appendix A) yielded an R_c value of .289104 with an associated *p* calculated value of .0376.

INSERT TABLE 2 ABOUT HERE

Table 3 presents an analysis illustrating the equivalence of *t*-tests and canonical correlation analysis. The *p* calculated value associated with the test of differences between means on variable CLASSAVG across the variable SEX, groups "1" and "2", was .7526. A canonical correlation analysis yielded a R_c value of .044781 with an associated *p* calculated value of .7526.

INSERT TABLE 3 ABOUT HERE

Table 4 presents a 2 x 2 factorial ANOVA for scores on the dependent variable SURVEY across ways defined by variables SEX and STATUS. Table 5 presents results from four separate canonical correlation analyses using different combinations of the *a priori* contrasts expressing the group membership information involved in the variables SEX and STATUS. The error effect for the full ANOVA model reported in Table 4 was .948948746 and may be obtained by dividing the sum of squares error by the sum of squares total (8523.38461538 / 8981.92307692).

INSERT TABLES 4 AND 5 ABOUT HERE

Table 6 presents lambda values associated with each of the canonical analyses analogous to each ANOVA model. Note that the lambda value associated with the full model was .9489487, the same value as the error effect for the full ANOVA model. Thompson (1988a) reported that 1 - lambda equals the squared canonical correlation coefficient. In this example, 1 - .9489487 = .051051, which is the squared canonical correlation for the canonical analysis and also the correlation ratio (eta^2) for the full ANOVA model. Thompson (1988a) explains that the multivariate lambda is analogous to the univariate sum of squares error divided by the sum of squares total.

INSERT TABLE 6 ABOUT HERE

Table 7 converts the canonical lambda's into separate effects for each ANOVA omnibus effect. Smaller lambda's indicate larger

effect sizes (1 - lambda = effect size). Note that in Table 4, the interaction of SEX*STATUS has the largest effect size .04 (315.077/8981.923). Now note that in Table 6 SEX*STATUS has the smallest lambda (.964351613) and that 1 - .964351613 = .04, which is also the ANOVA effect size for the interaction of SEX*STATUS. Thus, the smallest lambda reported in Table 6 is associated with the main effect for the interaction of SEX*STATUS in Table 4. Table 7 converts the Table 6 omnibus effect lambda's into ANOVA F tests comparable to those presented in Table 4.

INSERT TABLE 7 ABOUT HERE

Table 8 presents the multiple regression analysis in which variables AGE and GPA are used to predict the dependent variable SURVEY. Table 8 also contains the canonical correlation analysis involving the same variables for the prediction of the dependent variable SURVEY. Note that the results from both analyses are comparable except that canonical results are presented to more digits to the right of the decimal. Table 9 illustrates how canonical function coefficients are converted to the beta weights in multiple regression and vice versa (Thompson & Borrello, 1985).

INSERT TABLES 8 AND 9 ABOUT HERE

Table 10 presents a 2 X 2 factorial MANOVA involving CLASSAVG and SURVEY as the dependent variables and SEX and STATUS as the classification variables. Four separate canonical analyses are presented in Table 11. The canonical analyses were performed using

the contrast coded variables SX1, ST1, and SX1ST1, which respectively express the classification variables SEX and STATUS and the interaction of SEX and STATUS. Table 12 presents the conversion of the lambda values for the canonical analyses reported in Table 11 to comparable lambda values associated with the omnibus MANOVA effects (i.e., lambda values) shown in Table 10. Note that the lambda values in Table 10 and Table 12 are comparable.

INSERT TABLES 10 THRU 12 ABOUT HERE

Table 13 contains results from a discriminant analysis involving variables CLASSAVG and SURVEY to predict membership in the two groups, freshmen and sophomore, conveyed by the variable STATUS. Table 14 presents results from the canonical analysis involving the variables CLASSAVG and SURVEY as predictor variables and ST1 as the criterion variable. ST1 is the contrast coded variable which expresses the same information contained in the variable STATUS. Note that the results of Table 13 and Table 14 are directly comparable.

INSERT TABLES 13 AND 14 ABOUT HERE

Table 15 presents the conversion of function coefficients for the variables CLASSAVG and SURVEY for both the canonical and discriminant analysis. So that the canonical and discriminant results may be comparable, the largest function coefficient in each function is set equal to unity, or 1.00. This is done for both the canonical and discriminant analysis. Note that in the

canonical analysis, the function coefficient are .7447 for CLASSAVG and .6021 for SURVEY. To perform the conversions, the smaller coefficient, .6021, is divided by the larger coefficient, .7447, and the largest coefficient is set equal to 1.00 (.7447/.7447). The conversion is replicated for the discriminant analysis, as shown on Table 15.

INSERT TABLE 15 ABOUT HERE

Summary

It has been noted in the present paper that all parametric methods are special cases of canonical correlation analysis, and can in fact be performed using the canonical correlation procedures in commonly available computer statistics packages such as SAS. It was suggested that canonical analysis has (a) advantages of versatility, since the method can accomodate various levels of scale and does not require that variance be discarded or that variable relationships be distorted, (b) advantages of scientific parsimony, since the analysis is also a variable reduction technique not unlike principal components analysis, and (c) heuristic value, since the analytic method provides a framework to allow educational researchers to understand how all parametric methods are related. A small data set and the accompanying SAS program was presented, and it was established through concrete demonstration that canonical analysis does subsume other methods as special cases.

References

Afifi, A.A., & Clark, V. (1984). Computer-aided multivariate analysis. Belmont, CA: Lifetime Learning Publications.

Benton, R.L. (1990, January). The statistical power of planned comparisons. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.

Chacko, H.E. (1986, November). An example of the use of canonical correlation analysis. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis. (ERIC Document Reproduction Service No. ED 280 897)

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Fish, L. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counselling and Development, 21, 130-137.

Huberty, C.J., & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analysis. Psychological Bulletin, 105, 302-308.

Huberty, C.J., & Wisenbaker, J.M. (in press). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2). Greenwich, CT: JAI Press.

Keppel, G., & Zedeck, S. (1989). Data analysis for research designs. New York: W.H. Freeman.

Knapp, T. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin,

85, 410-416.

McMurray, M.A. (1987, January). The concepts underlying structure coefficients, communality, and factor scores in the exploratory factor analytic case. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service No. ED 284 873)

Maxwell, S.E. (in press). ̃ cent developments in MANOVA applications. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2). Greenwich, CT: JAI Press.

Stevens, J. (1986). Applied multivariate statistics for the social sciences. Hillsdale, NJ: Erlbaum.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Beverly Hills: SAGE.

Thompson, B. (1986, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-South Educational research Association, Memphis.

Thompson, B. (1988a, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1988b, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1989, August). Application of multivariate statistics: A bibliography of canonical correlation analysis studies. Paper presented at the annual meeting of the American Psychological Association, New Orleans. (ERIC Document Reproduction Service No. forthcoming).

Thompson, B., & Borrello, G.M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Table 1
Data from Mcmurray (1987)

01	10	10	06	08	03	10	06	06	10	08	1	1	25	2.0	75	1	01	01	01
02	10	06	07	10	09	09	08	09	07	06	2	1	27	2.0	85	1	-1	01	-1
03	09	04	02	10	08	08	06	09	10	03	1	2	22	2.6	87	1	01	-1	-1
04	09	07	06	06	08	08	08	07	07	06	2	2	29	1.6	72	1	-1	-1	01
05	09	10	01	07	06	01	06	07	10	04	1	1	20	1.9	65	1	01	01	01
06	10	10	06	10	10	08	10	08	08	09	2	1	27	2.7	88	1	-1	01	-1
07	10	10	02	02	10	10	10	10	10	10	1	2	26	1.2	44	1	01	-1	-1
08	08	10	01	08	09	07	08	09	09	06	1	1	27	3.4	91	1	01	01	01
09	10	08	08	06	08	09	08	08	08	08	2	2	22	4.0	98	1	-1	-1	01
10	09	10	08	09	10	09	10	09	08	10	2	2	21	2.3	82	1	-1	-1	01
11	10	07	07	05	10	10	09	08	08	06	2	1	25	4.0	99	1	-1	01	-1
12	10	09	04	10	06	03	07	10	10	01	1	1	22	2.4	85	1	01	01	01
13	06	06	02	07	05	02	03	04	06	02	2	2	23	2.2	79	1	-1	-1	01
14	10	19	04	10	10	08	10	09	10	07	2	1	21	2.7	89	1	-1	01	-1
15	09	07	05	02	08	05	08	06	08	08	1	2	27	4.0	95	1	01	-1	-1
16	08	08	04	02	04	06	05	02	08	04	2	2	29	1.6	69	1	-1	-1	01
17	09	09	05	04	10	08	08	05	05	09	1	2	20	3.3	88	1	01	-1	-1
18	02	05	03	08	05	09	06	08	07	05	1	1	20	1.7	72	1	01	01	01
19	10	09	02	08	10	06	10	07	10	10	2	1	29	1.5	55	1	-1	01	-1
20	10	09	07	02	09	08	09	05	06	07	1	2	27	3.4	91	1	01	-1	-1
21	10	10	10	10	10	10	10	10	10	10	2	1	19	2.9	90	1	-1	01	-1
22	10	10	10	10	10	10	10	02	10	10	2	1	25	3.9	94	1	-1	01	-1
23	09	10	02	05	09	05	09	07	08	10	1	2	22	3.4	92	1	01	-1	-1
24	10	09	04	08	07	01	07	09	10	01	2	1	23	3.7	97	1	-1	01	-1
25	10	10	01	08	08	08	05	07	10	05	1	2	23	2.9	88	1	01	-1	-1
26	10	10	10	10	10	09	09	10	10	10	1	1	20	1.8	74	1	01	01	01
27	10	10	05	05	05	08	08	07	10	05	2	2	29	3.7	90	1	-1	-1	01
28	10	06	07	08	08	10	09	08	09	07	2	2	23	1.8	60	1	-1	-1	01
29	10	10	10	01	08	10	08	02	02	05	2	1	18	2.1	75	2	-1	01	-1
30	03	06	03	04	08	10	09	06	05	06	1	1	19	3.8	99	2	01	01	01
31	10	10	06	10	08	10	10	02	08	10	1	1	26	2.7	89	2	01	01	01
32	09	09	05	09	09	10	08	09	09	10	2	2	21	3.3	94	2	-1	-1	01
33	06	07	04	04	09	03	09	05	07	08	2	1	21	1.6	59	2	-1	01	-1
34	10	10	01	04	07	05	01	04	08	01	2	2	20	3.5	94	2	-1	-1	01
35	10	08	05	04	08	09	08	06	07	09	1	2	22	3.3	92	2	01	-1	-1
36	07	05	05	04	09	10	06	03	05	07	2	2	26	1.6	54	2	-1	-1	01
37	06	08	10	08	08	10	08	08	08	08	1	1	29	2.2	71	2	01	01	01
38	08	09	08	09	09	08	08	07	09	06	1	2	27	1.5	48	2	01	-1	-1
39	08	09	05	09	09	08	03	08	08	08	2	1	21	2.4	83	2	-1	01	-1
40	07	06	09	06	07	09	05	03	07	06	1	2	22	2.4	33	2	01	-1	-1
41	05	08	01	02	05	10	02	02	04	01	2	1	20	2.6	85	2	-1	01	-1
42	10	06	10	01	09	10	10	05	05	10	1	2	25	1.7	65	2	01	-1	-1
43	10	10	05	07	08	06	08	06	09	08	2	2	20	3.2	85	2	-1	-1	01
44	05	04	02	08	07	02	06	08	09	03	1	1	25	3.5	95	2	01	01	01
45	09	09	03	10	06	09	04	08	08	09	2	2	25	2.4	78	2	-1	-1	01

Table 1 (continued)

46	07	10	05	08	09	03	08	07	10	07	1	2	30	2.8	83	2	01	-1	-1
47	05	10	04	09	03	09	04	04	10	04	1	1	24	2.2	80	2	01	01	01
48	08	06	01	06	10	01	09	07	05	05	1	1	30	1.0	52	2	01	01	01
49	06	08	03	06	07	07	07	06	06	05	2	2	26	1.8	63	2	-1	-1	01
50	01	09	05	09	05	06	02	08	08	06	1	2	24	1.2	64	2	01	-1	-1
51	10	10	02	09	06	04	05	09	10	04	2	1	27	2.4	75	2	-1	01	-1
52	08	07	04	03	08	08	09	03	06	08	1	1	25	3.5	91	2	01	01	01

Table 2
CCA Subsumes Pearson r [CLASSAVG with AGE]

Canonical Analysis		Pearson r	
Squared Rc	.083581		
Rc	.289104	r	-.28910
lambda	.9164187		
F	4.5602		
df	1/10		
p calc	.0376	p calc	.0376

Note. Unlike *r*, R_c can never be negative, but the magnitude of the two coefficients will always be equal.

Table 3
CCA Subsumes *t*-tests [CLASSAVG by SEX (1,2)]

Canonical Analysis		t-test analysis	
Squared Rc	.002005	Mean Group 1	79.19230769
Rc	.044781	SD	15.18161844
lambda	.9979947	Mean Group 2	80.46153846
		SD	13.65351462
F	.1005	t	-.3170
df	1/50	df	50
p calc	.7526	p calc	.7526

Table 4
Factorial ANOVA [SURVEY by SEX and STATUS]

Source	SOS	df	MS (SOS/df)	F	p calc
SEX	78.76923077	1	[78.76923077]	.44	.5086
STATUS	64.69230769	1	[64.69230769]	.36	.5490
SEX*STATUS	315.07692308	1	[315.07692308]	1.77	.1891
Error	8523.38461538	48	[177.5705128]		
Total	8981.92307692	51	[176.1161387]		

Table 5
Canonical Analyses Using Four Models

Model	Predictors of SURVEY			lambda
1	SX1	ST1	SX1ST1	.9489487
2	SX1		SX1ST1	.9561512
3		ST1	SX1ST1	.9577185
4	SX1	ST1		.9840277

Table 6
Conversion to ANOVA lambda's

Effect	Models	Conversion	lambda
SEX	1/3	.9489487/.9577185	.990843029
STATUS	1/2	.9489487/.9561512	.992467195
SEX*STATUS	1/4	.9489487/.9840277	.964351613

Table 7
Conversion of lambda's to ANOVA's F's

Source	(1-lambda / lambda)	X (df error/df effect)		= F Calc
SEX	1-.990843029/.990843029	* 48	/ 1	= .44
STATUS	1-.992467195/.992467195	* 48	/ 1	= .36
SEX*STATUS	1-.964351613/.964351613	* 48	/ 1	= 1.77

Table 8
CCA Subsumes Multiple Correlation [SURVEY with AGE and GPA]

Canonical Analysis		Regression Analysis	
Squared Rc	.010772	Squared R	.0108
Rc	.103787		
lambda	.9892282		
F	.2668	F	.267
df	2/49	df	2/49
p calc	.7669	p calc	.7669

Table 9
Function Coefficient and Beta Weight Conversions

Predictor	Function Coefficient	* Rc(or R)	= Beta Weight/Rc(or R)	= Function Coefficient
AGE	.3684	* .103787	.038236/.103787	= .3684
GPA	1.0025	* .103787	.104045/.103787	= 1.0024

Table 10
Factorial MANOVA [CLASSAVG, SURVEY by SEX, STATUS]

Source	lambda	F	df	p calc
SEX	.98937049	.25	2,47	.7779
STATUS	.98330095	.40	2,47	.6732
SEX*STATUS	.96302619	.90	2,47	.4126

Table 11
Canonical Analyses Using Four Models

Model	Predictors of CLASSAVG,SURVEY			lambda
1	SX1	ST1	SX1ST1	.9377301
2	SX1		SX1ST1	.9536552
3	ST1		SX1ST1	.9478048
4	SX1	ST1		.9737327

Table 12
Conversion to MANOVA lambda's

Effect	Models	Conversion	Result
SEX	1/3	.9377301/.9478048	.989370490
STATUS	1/2	.9377301/.9536552	.983300987
SEX*STATUS	1/4	.9377301/.9737327	.963026198

Table 13
CCA Subsumes Discriminant Analysis [STATUS with CLASSAVG,SURVEY]

Discriminant Analysis Results

	Function I
Squared Rc	.016179
Rc	.127197
lambda	.983821
F	.403
df	2/49
p calc	.6706

Table 14
CCA Subsumes Discriminant Analysis [ST1 with CLASSAVG,SURVEY]

Canonical Analysis Results

	Function I
Squared Rc	.016179
Rc	.127197
lambda	.983821
F	.403
df	2/49
p calc	.6706

Table 15
Conversion of Function Coefficients for Comparison

Canonical Correlation Analysis

	Function I	Result
SURVEY	.6021 / .7477 =	.8036
CLASSAVG	.7477 / .7477 =	1.0000

Discriminant Function Analysis

	Function I	Result
SURVEY	.6011 / .7464 =	.8053
CLASSAVG	.7464 / .7464 =	1.0000

APPENDIX A:
SAS Log File Used to Execute Example Analyses with Table 1 Data

```
DATA CCA;
INFILE MCMURRAY;
INPUT ID 1-2 V1 4-5 V2 7-8 V3 10-11 V4 13-14 V5 16-17 V6 19-20
V7 22-23 V8 25-26 V9 28-29 V10 31-32 SEX 34 STATUS 36 AGE 38-39
GPA 41-43 CLASSAVG 45-46 INVARGP 48 SX1 50-51 ST1 53-54
SX1ST1 56-57;
SURVEY=SUM(OF V1-V10);
PROC PRINT;
VAR V1-V10  SURVEY  SEX  STATUS  SX1  ST1  SX1ST1  AGE  GPA  CLASSAVG
INVARGP;
TITLE 'DESCRIPTION OF RAW DATA';
PROC MEANS;
VAR V1-V10 SURVEY SEX STATUS AGE GPA CLASSAVG;
TITLE 'DESCRIPTIVE STATISTICS FOR ALL PARTICIPANTS';
PROC CORR;
VAR V1-V10 SURVEY;
TITLE 'CORRELATION OF SURVEY ITEMS';
PROC CORR;
VAR SEX STATUS AGE GPA CLASSAVG SURVEY;
TITLE 'CORRELATION OF PREDICTOR AND CRITERION VARIABLES';
PROC CANCORR ALL;
VAR CLASSAVG;
WITH AGE;
TITLE 'CCA SUBSUMES PEARSON CORRELATION';
PROC CORR;
VAR AGE CLASSAVG;
TITLE 'CCA SUBSUMES PEARSON CORRELATION';
PROC CANCORR ALL;
VAR CLASSAVG;
WITH SEX;
TITLE 'CCA SUBSUMES T-TESTS';
PROC TTEST;
CLASS SEX; VAR CLASSAVG;
TITLE 'CCA SUBSUMES T-TESTS';
PROC CANCORR ALL;
VAR SURVEY; WITH SX1 ST1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR ALL;
VAR SURVEY; WITH SX1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR ALL;
VAR SURVEY; WITH ST1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR ALL;
VAR SURVEY; WITH SX1 ST1;
TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC ANOVA;
CLASS SEX STATUS;
```

```
MODEL SURVEY=SEX STATUS SEX*STATUS;
TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR ALL;
VAR SURVEY;
WITH AGE GPA;
TITLE 'CCA SUBSUMES MULTIPLE REGRESSION';
PROC REG;
MODEL SURVEY=AGE GPA/STB;
TITLE 'CCA SUBSUMES MULTIPLE REGRESSION';
PROC CANCORR ALL;
VAR CLASSAVG SURVEY; WITH SX1 ST1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR ALL;
VAR CLASSAVG SURVEY; WITH SX1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR ALL;
VAR CLASSAVG SURVEY; WITH ST1 SX1ST1;
TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR ALL;
VAR CLASSAVG SURVEY; WITH SX1 ST1;
TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC ANOVA;
CLASS SEX STATUS;
MODEL SURVEY CLASSAVG=SEX STATUS SEX*STATUS;
MANOVA H=_ALL_/SUMMARY;
TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR ALL;
VAR ST1;
WITH CLASSAVG SURVEY;
TITLE 'CCA SUBSUMES DISCRIMINANT ANALYSIS';
PROC CANDISC ALL;
VAR CLASSAVG SURVEY;
CLASS STATUS;
TITLE 'CCA SUBSUMES DISCRIMINANT ANALYSIS';
```